Texture and Color-based Image Retrieval Using the Local Extrema Features and Riemannian Distance

Minh-Tan Pham, Grégoire Mercier, Lionel Bombrun, and Julien Michel

***Abstract*—A novel efficient method for content-based image retrieval (CBIR) is developed in this paper using both texture and color features. Our motivation is to represent and characterize an input image by a set of local descriptors extracted at characteristic points (i.e. keypoints) within the image. Then, dissimilarity measure between images is calculated based on the geometric distance between the topological feature spaces (i.e. manifolds) formed by the sets of local descriptors generated from these images. In this work, we propose to extract and use the local extrema pixels as our feature points. Then, the so-called local extrema-based descriptor (LED) is generated for each keypoint by integrating all color, spatial as well as gradient information captured by a set of its nearest local extrema. Hence, each image is encoded by a LED feature point cloud and riemannian distances between these point clouds enable us to tackle CBIR. Experiments performed on Vistex, Stex and colored Brodatz texture databases using the proposed approach provide very efficient and competitive results compared to the state-of-the-art methods.**

***Index Terms*—Content-based image retrieval (CBIR), local textures, pointwise approach, local extrema features, riemannian distance (RD).**

I. Introduction

CONTENT-BASED image retrieval (CBIR) has been so far tackled by many research studies in the literature. Texture and color, together with other features such as shape, edge, surface, etc., play significant roles in CBIR systems thanks to their capacity of providing important parameters and characteristics not only for human vision but also for automatic computer-based visual recognition. Texture has appeared in most of CBIR frameworks while color has been more and more exploited to improve retrieval performance, in particular for natural images.

Some recent comprehensive surveys on the CBIR field can be found in [1]–[4]. Among state-of-the-art propositions, a great number of multiscale texture representation and analysis methods using probabilistic approach have been developed within the past two decades. In [5], the authors propose to model the spatial dependence of pyramidal discrete wavelet transform (DWT) coefficients using the generalized Gaussian distributions (GGD) and the dissimilarity measure between images is derived based on the Kullback-Leibler divergences (KLD) between GGD models. Sharing the similar principle, multiscale coefficients yielded by the discrete cosine transform (DCT), the dual-tree complex wavelet transform (DT-CWT) or its rotated version (DT-RCWT) are modeled by different statistical models such as GGD, the multivariate Gaussian mixture models (MGMM), Gaussian copula (GC), Student-t copula (StC), or other distributions like Gamma, Rayleigh, Weibull, Laplace, etc. to perform texture-based image retrieval [6]–[10]. Recently, the Gabor wavelets [11] and the normalized scattering window transform (NSWT) [12] have been proposed to enrich this multiscale analysis approach with some improvements in retrieval accuracy compared against previous schemes. Last but not least, by taking into account color information within these probabilistic approaches, several studies have provided significant improvement for color image retrieval [13], [14].

The second family of methods which has drawn attention of many researchers and has provided quite effective CBIR performance is the local pattern-based framework. Local binary patterns (LBP) descriptor, which compares neighboring pixels to the center pixel and affects them to 0 and 1 to form a binary number, was first embedded in a multiresolution and rotation invariant scheme for texture classification in [15]. Inspired from this work, many propositions have been developed for texture retrieval and classification such as the local derivative patterns (LDP) [16], the local maximum edge binary patterns (LMEBP) [17], the local ternary patterns (LTP) [18], the local tetra patterns (LTrP) [19], etc. These descriptors, particularly the LTrP, provide quite efficient texture retrieval performance. However, due to the fact that they are applied to gray-scale images, their performance on natural images is limited without exploiting color information. To overcome this issue, several recent strategies have been proposed to incorporate these local patterns with color features. Some techniques can be stated here are the joint histogram of color and local extrema patterns (LEP + colorhist) [20], the local oppugnant color texture pattern (LOCTP) [21] and the local extrema co-occurrence pattern (LECoP) [22].

Another CBIR system which has offered very competitive results within the past few years relies on an image compression approach called the block truncation coding (BTC). The first BTC-based retrieval scheme for color images was proposed in [23] followed by some improvements during a few years later [24], [25]. Until very recently, one has been witnessing the evolution of BTC-based retrieval frameworks such as the ordered-dither BTC (ODBTC) [26], [27], the error diffusion BTC (EDBTC) [28] and the dot-diffused BTC (DDBTC) [29]. Within these approaches, an image is divided into multiple non-overlapping blocks and one of the BTC-based systems compresses each block into the so-called color quantizer and bitmap image. Then, to characterize the image content, a feature descriptor is constructed using the color histogram feature (CHF) or the color co-occurrence feature (CCF) combined with the bit pattern feature (BPF) which describes edge, shape and texture information of the image.

These features are extracted from the above color quantizer and bitmap image. To the best of our knowledge, along with the previously-mentioned LECoP method [22], these BTC-based techniques [26]–[29] currently provide the most efficient performance on texture and color-based CBIR task.

In this work, we would like to develop a powerful texture and color image retrieval strategy based on the capacity of characteristic points to capture significant information from the image content. Due to the fact that natural images usually involve a variety of local textures and structures which do not appear homogeneous within the entire image, an approach taking into account local features could become relevant. That may be the reason why most local feature-based CBIR schemes (e.g. LTrP [19], LOCTP [21], LECoP [22], etc.) or BTC-based approaches [26]–[29] (i.e. which in fact sub-divide each query image into multiple blocks) have achieved better retrieval performance than probabilistic methods which model the entire image using different statistical distributions [5]–[14]. We will provide later their performance for a comparison within our experimental study. Back to this paper, by taking into consideration the above-mentioned remark, our motivation here is to represent and characterize an input image by a set of local descriptors generated only for characteristic points detected from the image. This local descriptor set can be considered as a feature point cloud characterizing the image. Hence, an appropriate measure between these point clouds enables us to derive the dissimilarity between images for retrieval task.

Indeed, characteristic points (i.e. keypoints) to be detected are required to be able to encapsulate important textural and contextual information within the image content. Recent propositions [30], [31] addressing texture-based image segmentation and classification in the scope of very high resolution (VHR) remote sensing imagery have proved the capacity of the local extrema (i.e. local maximum and local minimum pixels in terms of intensity) to capture and represent textural features from an image. Their performance on texture representation and discrimination using a pointwise approach (i.e. non-dense approach based on characteristic points) has been confirmed and validated for this kind of images, in particular to tackle local textures with a weakly-verified (or without) stationarity hypothesis [30]. In this work, we continue and improve the idea of using the local extrema pixels to tackle CBIR. By embedding both color and geometric information, as well as gradient features, captured by these feature points, we propose the local extrema-based descriptor (LED) for texture and color description. As a result, an input image will be encoded by a LED feature point cloud. Moreover, we propose to exploit a geometric-based distance measure, i.e. the riemannian distance (RD) [32] between the feature covariance matrices of these point clouds, for dissimilarity measurement. This distance takes into account the topological structure of each point cloud within the LED feature space. Therefore, it becomes an effective measure of dissimilarity in our CBIR scheme.

The remainder of this paper is organized as follows. Section II presents the pointwise approach for local texture representation and description using local extrema features and the construction of LED descriptors. The proposed CBIR framework is described in details in Section III. Section IV carries out our experimental study and provides comparative results of the proposed algorithm against existing methods. A conclusion and some perspectives of our work are discussed in Section V.

II. Local texture representation and description using local extrema features

A. Approach

The idea of using the local maximum and local minimum pixels for texture representation and characterization has been introduced in [30], [31] in the scope of VHR optical satellite images. Regarding to this point of view, an image texture is formed by a certain spatial arrangement of pixels holding some variations of intensity. Hence, different textures are reflected by different types of pixel's spatial arrangements and intensity variations. These meaningful properties can be approximated by the local maximum and local minimum pixels extracted from the image. These local extrema have been proved to be able to capture the important geometric and radiometric information of the image content, hence relevant for texture analysis and description. In this work, we exploit and improve the capacity of local extrema pixels to characterize both texture and color features within CBIR context. Let us first recall their definition and extraction.

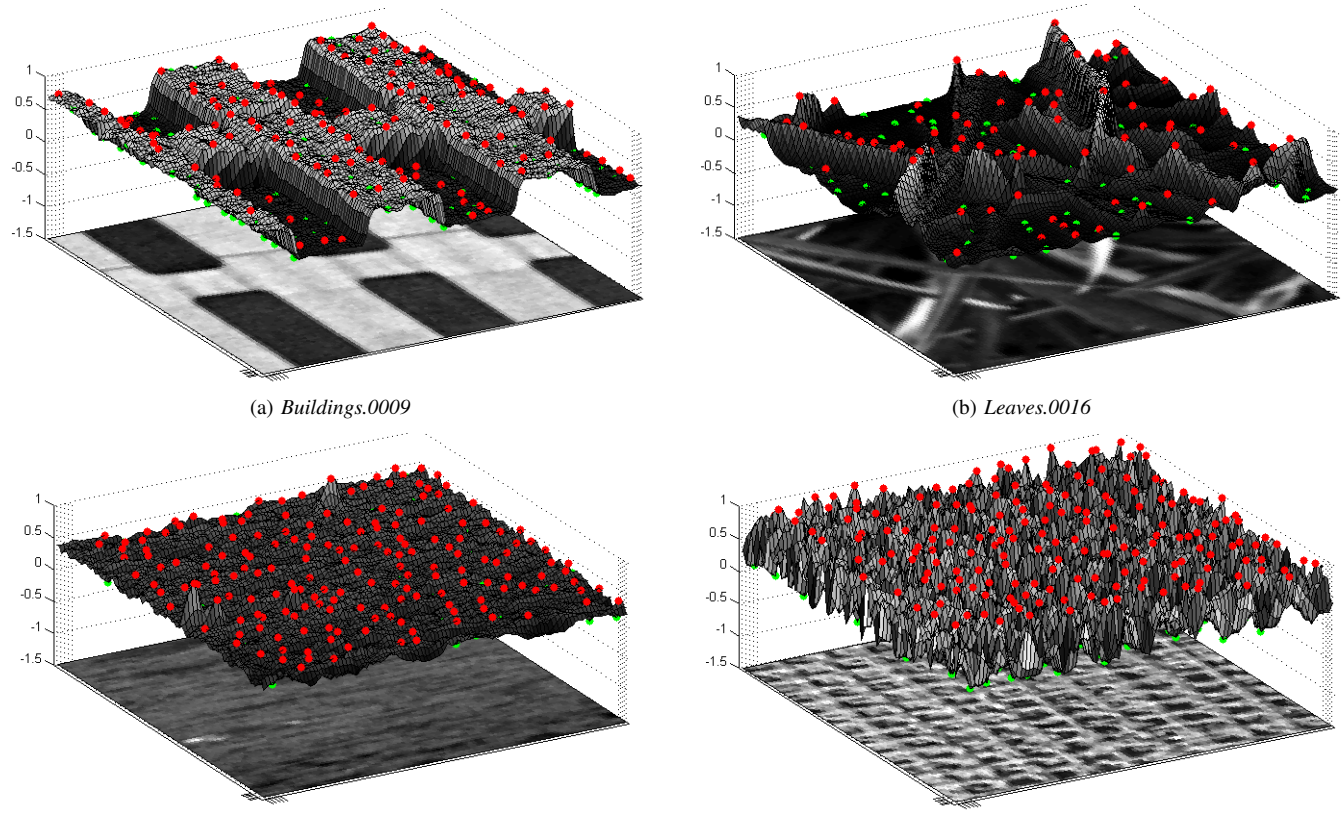
A pixel in a grayscale image is supposed to be a local maximum (resp. local minimum) if it holds the highest (resp. lowest) intensity value within a neighborhood window centered at it. Let $S_\omega^{\max}(I)$ and $S_\omega^{\min}(I)$ denote the local maximum and local minimum sets extracted from a grayscale image I using the $\omega \times \omega$ search window. Let $p = (x_p, y_p)$ be a pixel located at position (x_p, y_p) on the image plane having its intensity value $I(p)$, we have:

$$p \in S_\omega^{\max}(I) \Leftrightarrow \left\{ I(p) = \max_{q \in \mathcal{N}_{\omega\times\omega}(p)} I(q) \right\}, \quad (1)$$

$$p \in S_\omega^{\min}(I) \Leftrightarrow \left\{ I(p) = \min_{q \in \mathcal{N}_{\omega\times\omega}(p)} I(q) \right\}, \quad (2)$$

where $\mathcal{N}_{\omega\times\omega}(p)$ represents a set of pixels inside the $\omega \times \omega$ neighborhood window of p. We note that there are many ways to extend the above definition for color images (i.e. detecting on the grayscale version, using the union or intersection of subsets detected on each color channel, etc.). In this work, we propose to extract the local extrema pixels from the grayscale version of color image since this produced superior performance within most of our experiments.

Fig. 1 illustrates the capacity of the local max and min pixels to represent, characterize and discriminate different textures which are extracted from the Vistex image database [33]. Each image patch consists of 100×100 pixels. We display for each one a 3-D surface model using the image intensity as the surface height. The local max pixels (in red) and local min pixels (in green) are extracted following (1) and (2) by a 5×5 search window. Some green points may be unseen since they are obscured by the surface. We observe from the figures how these local extrema appear within each texture. Their



(a) *Buildings.0009* (b) *Leaves.0016*

(c) *Water.0005* (d) *Fabric.0017*

Fig. 1. Illustration: spatial distribution and arrangement of local maximum pixels (in red) and local minimum pixels (in green) within 4 different textures (size of 100×100 pixels) extracted from the Vistex database [33]: *Buildings.0009*, *Leaves.0016*, *Water.0005* and *Fabric.0017*. The local extrema are detected using a 5×5 search window. The figure is better visualized in colors.

spatial information (relative distance, direction, density, etc.) and intensities can be encoded to describe and discriminate each texture, which leads to the pointwise approach of this paper.

Last but not least, one may wonder about the possibility of using other popular keypoints (Harris corners, Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), etc. [34]) to perform our strategy. In fact, keypoints in this work are used to represent the image, but also to describe textural features. Hence, they are expected to be able to capture all image's contextual information and to appear within all texture zones. Local extrema pixels satisfy this requirement since they are extracted within any variation of intensity (i.e. appearance of textures). Meanwhile, such keypoints like Harris, SIFT or SURF are normally focused on corners, edges, and salient features so that they may not probably be detected from some quite homogeneous textures within natural scenes such as sand, water, early grass fields, etc. Hence, they are not relevant for our approach which describes textural features using only characteristic points.

B. Generation of Local Extrema-based Descriptor (LED)

As previously discussed, our work not only considers the local extrema pixels as characteristic points to represent the image but also exploits them to construct texture and color descriptors. The local extrema-based descriptor (LED) is generated for each keypoint by integrating the color, spatial and gradient features captured by its nearest local maxima and local minima on the image plane. Two strategies can be considered for the research of nearest local extrema around each keypoint:

1) Fix the number (N) of nearest local maxima and nearest local minima for each one.
2) Or, fix a window size $W \times W$ around each keypoint, then all local maxima and minima inside that window are considered.

The first strategy takes into account the local properties around keypoints since the implicit neighborhood size considered for each one varies depending on the density of local extrema around it (i.e. the search process goes further to look for enough extrema when their density is sparse). On the other hand, by fixing the window size, the second approach considers equivalent contributions of neighboring environments for all keypoints and hence better deals with outlier points. Moreover, its implementation is less costly than the first one, especially for large-size image (i.e. the search of N points is more costly within large-size image). Nevertheless, our experimentation shows that both strategies provide similar performance for the studied data sets. In this paper, we describe the approach considering a fixed window size for all keypoints, but it is worth noting that a similar principle can be applied to build the proposed descriptor using the first approach.

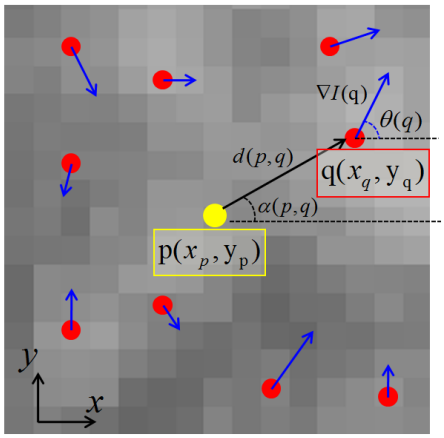


Fig. 2. Geometric and gradient information from a local maximum (resp. local minimum) pixel $q = (x_q, y_q)$ within $\mathcal{N}_W^{\max}(p)$ (resp. $\mathcal{N}_W^{\min}(p)$) considered for the calculation of LED descriptor at the studied keypoint $p = (x_p, y_p)$. Here, $d(p,q)$ is the distance between p and q; $\alpha(p,q)$ is the angle of the vector yielded from p to q. We have $d(p,q) = d(q,p)$ but $\alpha(p,q) \neq \alpha(q,p)$. Then, $\nabla I(q)$, $\theta(q)$ are the gradient magnitude and gradient orientation at q.

The generation of LED descriptors is described as follows. Let $\mathcal{N}_W^{\max}(p)$ and $\mathcal{N}_W^{\min}(p)$ be the two set of local maximum and local minimum pixels inside the $W \times W$ window around the understudied keypoint p, the following features involving color, spatial and gradient information are extracted for each set. For better explanation, we display in Fig. 2 the geometric and gradient features derived from each local max or local min within these sets. Below are the features extracted from $\mathcal{N}_W^{\max}(p)$, the feature generation for $\mathcal{N}_W^{\min}(p)$ is similar.

1) Mean and variance of three color channels:

$$\mu_c^{\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} I_c(q), \tag{3}$$

$$\sigma_c^{2\,\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} (I_c(q) - \mu_c^{\max}(p))^2, \tag{4}$$

where $c \in \{\text{red}, \text{green}, \text{blue}\}$ represents the three color components and $|\mathcal{N}|$ is the cardinality of the set $\mathcal{N}$.

2) Mean and variance of spatial distances from each local maximum to point p:

$$\mu_d^{\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} d(p,q), \tag{5}$$

$$\sigma_d^{2\,\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} (d(p,q) - \mu_d^{\max}(p))^2, \tag{6}$$

where $d(p,q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2}$ is the spatial distance between two points p and q on the image plane. We remind that $p = (x_p, y_p)$ and $q = (x_q, y_q)$.

3) Circular variance [35] of angles of geometric vectors formed by each local maximum and point p:

$$\sigma_{\text{cir},\alpha}^{2\,\max}(p) = 1 - \sqrt{\bar{c}_\alpha(p)^2 + \bar{s}_\alpha(p)^2}, \tag{7}$$

where

$$\bar{c}_\alpha(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \cos\alpha(p,q),$$

$$\bar{s}_\alpha(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \sin\alpha(p,q),$$

$$\alpha(p,q) = \arctan\left(\frac{y_q - y_p}{x_q - x_p}\right), \ \alpha(p,q) \in [-\pi, \pi], \forall p, q.$$

4) Mean and variance of gradient magnitudes:

$$\mu_g^{\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \nabla I(q), \tag{8}$$

$$\sigma_g^{2\,\max} = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \left(\nabla I(q) - \mu_g^{\max}(p)\right)^2, \tag{9}$$

where ∇I is the gradient magnitude image obtained by applying the Sobel filter on the gray-scale version of the image.

5) Circular variance [35] of gradient orientations:

$$\sigma_{\text{cir},\theta}^{2\,\max}(p) = 1 - \sqrt{\bar{c}_\theta(p)^2 + \bar{s}_\theta(p)^2}, \tag{10}$$

where

$$\bar{c}_\theta(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \cos\theta(q),$$

$$\bar{s}_\theta(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \sin\theta(q),$$

θ is the gradient orientation image obtained together with ∇I by Sobel filtering.

All of these features are then integrated into the feature vector $\delta^{\max}(p)$, which encodes the local max set around p:

$$\begin{aligned}\delta^{\max}(p) = \Big[&\mu_{\text{red}}^{\max}(p), \sigma_{\text{red}}^{2\,\max}(p), \mu_{\text{green}}^{\max}(p), \sigma_{\text{green}}^{2\,\max}(p),\\ &\mu_{\text{blue}}^{\max}(p), \sigma_{\text{blue}}^{2\,\max}(p), \mu_d^{\max}(p), \sigma_d^{2\,\max}(p),\\ &\sigma_{\text{cir},\alpha}^{2\,\max}(p), \mu_g^{\max}(p), \sigma_g^{2\,\max}(p), \sigma_{\text{cir},\theta}^{2\,\max}(p)\Big].\end{aligned} \tag{11}$$

The generation of $\delta^{\min}(p)$ is similar. Now, let $\delta^{\text{LED}}(p)$ be the LED feature vector extracted for p, we have:

$$\delta^{\text{LED}}(p) = \left[I_{\text{red}}(p), I_{\text{green}}(p), I_{\text{blue}}(p), \delta^{\max}(p), \delta^{\min}(p)\right]. \tag{12}$$

The proposed descriptor $\delta^{\text{LED}}(p)$ enables us to characterize the local environment around p by understanding how local maxima and local minima are distributed and arranged, and also how they capture structural properties (given by gradient features) as well as color information. We note that LED descriptors are invariant to rotation. As observed from the feature computation process, for the two directional features including angles α and gradient orientations θ, only their circular variance [35] is taken into account, their circular mean is not involved to ensure the rotation-invariant property.

The feature dimensionality of δ^{LED} in (12) is equal to 27. One may realize that this descriptor can be feasibly improved by modifying or adding other features into the vector. One of

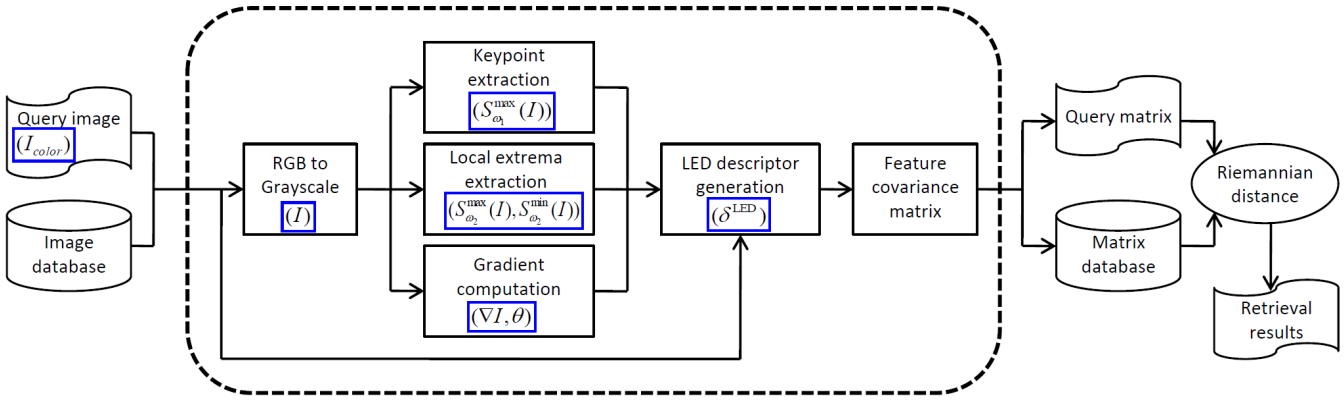


Fig. 3. Proposed framework for texture and color-based image retrieval using the local extrema-based descriptors (LED) and riemannian distance (RD).

the improvements can be considered here is to involve other filtering approaches for gradient computation. For example, beside the gradient images $(\nabla I, \theta)$, it is interesting to add $(\nabla I_\sigma, \theta_\sigma)$ which are generated by smoothing the image by a lowpass Gaussian filter ($I_\sigma = G_\sigma * I$) before applying the Sobel operator. Then, similar gradient features are extracted as in (8), (9) and (10). By inserting these features (computed for two extrema sets) into $\delta^{\text{LED}}(p)$, an enhanced descriptor of dimension 33 is built. Both 27-D and 33-D versions of LED descriptor are going to be experimented later in the paper.

III. Proposed Framework for Texture and Color Image Retrieval

The proposed retrieval algorithm consists of two primary stages: the extraction of LED texture and color descriptors to characterize each query image from the image database and the computation of distance measure for retrieval process. Each of them is now described in details. Then, the complete framework is presented.

A. LED Feature Extraction

The previous section has proved the capacity of the local extrema pixels to represent and characterize both local texture and color information from the image content. In this section, each query image will be encoded by a set of LED descriptors generated at characteristic points extracted from the image. As discussed in Section II-A, our favorite keypoints are also the local extrema points. One could exploit both local maxima and local minima as keypoints. Our experimentation shows that using one of the two can provide as good performance as employing both. Hence, in this article, we propose that our characteristic points are the local maximum pixels detected by a search window of size $\omega_1 \times \omega_1$. LED descriptors are then generated for each keypoint with the support of the local max and local min sets extracted by another search window of size $\omega_2 \times \omega_2$. Here, we separate the two window sizes since ω_2 should be small enough to ensure sufficient max and min pixels around each keypoint (i.e. dense enough) for descriptor computation. On the other hand, the density of keypoints within the image can be lower than or equal to the density of local extrema by setting $\omega_1 \geq \omega_2$. It is revealed that a coarser density of keypoints (by increasing ω_1) still produces good retrieval results but reduces pretty much computation time. The sensitivity of the proposed method to these window sizes will be analyzed in Section IV-C3.

Subsequently, a set of LED descriptors is constructed for the keypoint set to characterize the image. Given a neighborhood window size $W \times W$, the computation of LED feature vector for each keypoint is carried out as described in Section III-A (see Equations (3) to (12)). Similarly, the algorithm sensitivity to parameter W is discussed in Section IV-C3.

B. Dissimilarity Measure for Retrieval

Each query image from the database is now represented and characterized by a set of LED descriptors which can be considered as a LED feature point cloud. The dissimilarity measure between two images becomes the distance between two LED feature point clouds. Several metrics can be performed. One may consider that these sets of feature vectors follow a multivariate normal distribution and thus employ the simplified Mahalanobis distance or the symmetric Kullback-Leibler distance for dissimilarity measurement as follows:

$$d_{\text{Mahal}} = (\mu_1 - \mu_2)^T \left(C_1^{-1} + C_2^{-1}\right)(\mu_1 - \mu_2), \tag{13}$$

$$\begin{aligned} d_{\text{KL}} = {} & \text{trace}\left(C_1 C_2^{-1} + C_2 C_1^{-1}\right) \\ & + (\mu_1 - \mu_2)^T \left(C_1^{-1} + C_2^{-1}\right)(\mu_1 - \mu_2), \end{aligned} \tag{14}$$

where μ_1, C_1, μ_2, C_2 are the estimated means and covariance matrices of the normal distributions for the two point clouds.

Others would like to only account for the space of the feature covariance matrices by supposing that these point clouds have the same mean feature vector. Hence, their dissimilarity measure is calculated based on the distance between their covariance matrices. Our proposition in this paper is to utilize a geometric-based distance between the feature covariance matrices estimated from each set of LED descriptors. Thanks to the fact that these feature covariance matrices possess a positive semi-definite structure, their riemannian distance (which is also called geodesic distance in some studies) is proposed to become the distance measure in our retrieval

scheme. The riemannian distance between two covariance matrices is defined in [32] as follows:

$$d_{RD}(C_1, C_2) = \sqrt{\sum_{\ell=1}^{d} \log^2 \lambda_\ell} \tag{15}$$

where λ_ℓ is the ℓ^{th} generalized eigenvalue satisfying $\lambda_\ell C_1 \chi_\ell - C_2 \chi_\ell = 0, \ell = 1 \ldots d$. χ_ℓ is the corresponding eigenvector to λ_ℓ and d is the feature dimensionality of LED feature vectors (i.e. 27-D or 33-D as described in Section II-B).

Experiments show that the riemannian distance (RD) outperforms the Mahalanobis and the symmetric Kullback-Leibler distances in terms of both retrieval accuracy and computational time. Moreover, by using the RD, only one feature covariance matrix needs to be computed and stocked for each query image. On the other hand, the distances in (13) and (14) also involve the mean feature vector for their computation, hence require more computer's storage memory.

Last but not least, for better confirming and validating our choice of riemannian distance, we will provide in Section IV-C4 a detailed comparison (i.e. in terms of retrieval performance and computation time) of this metric not only with the mentioned Mahalanobis and Kullback-Leibler metrics but also against some other distance measures of covariance matrices such as the log-euclidean, the Wishart-like and the Bartlett distances [36], [37].

C. Proposed Retrieval Framework

The full proposed framework for texture and color-based image retrieval using LED features and riemannian distance is outlined in Fig. 3. The algorithm can be highlighted as follows:

1. Load the query color image I_{color}.
2. Convert the image to grayscale I.
3. Compute gradient images from I:
   + $(\nabla I, \theta)$ for the 27-D version,
   + $(\nabla I, \nabla I_\sigma, \theta, \theta_\sigma)$ for the enhanced 33-D version.
4. Extract the keypoint set and the two local extrema sets from I:
   + keypoint set: $S(I) = S^{\max}_{\omega_1}(I)$,
   + extrema sets: $S^{\max}_{\omega_2}(I)$ and $S^{\min}_{\omega_2}(I)$.
5. Generate LED descriptors for all keypoints:
   + $\forall p \in S(I)$, extract $\delta^{\text{LED}}(p)$ as (12).
6. Estimate the feature covariance matrice for these LED descriptors.
7. Compute the riemannian distance (15) between the query and the other images from the database.
8. Sort these distance measures and produce the best matches as the final retrieval result for the query.

IV. Experimental Study

A. Image databases

Three color texture image databases including the MIT Vision Texture database (Vistex) [33], the Salzburg Texture database (Stex) [38] and the novel Colored Brodatz Texture database (CBT) [39] are exploited to perform our experimental study.

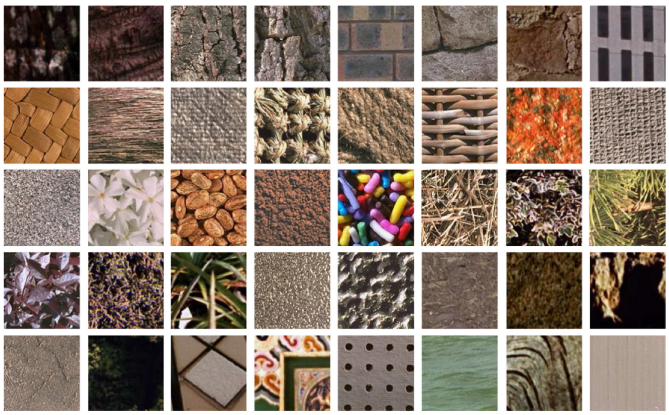


Fig. 4. 40 texture classes within the Vistex database [33].

Fig. 5. Some sample texture classes within the Stex database [38].

Vistex is one of the most widely used texture databases for performance assessment and comparative study in the CBIR field. It consists of 40 texture images of size 512×512 pixels. As in most literature studies, each texture is divided into 16 non-overlapping subimages of size 128×128 pixels in order to create a database of 640 images (i.e. 40 classes $\times$ 16 images/class). In the rest of the paper, the database is named Vistex-640. Being much larger, the Stex database is a collection of 476 color texture images captured in the area around Salzburg, Austria under real-word conditions. Like for Vistex, each 512×512 texture image is divided into 16 non-overlapping patches to build the so-called Stex-7616 database. Illustrations of the 40 Vistex textures and some Stex texture samples can be visualized in Figs. 4 and 5.

Next, the Colored Brodatz Texture (CBT) database is an extension of the gray-scale Brodatz texture database [39]. This database both preserves the rich textural content of the original Brodatz and possesses a wide variety of color content. Hence, it becomes relevant for the evaluation of texture and color-based CBIR algorithms. The CBT database consists of

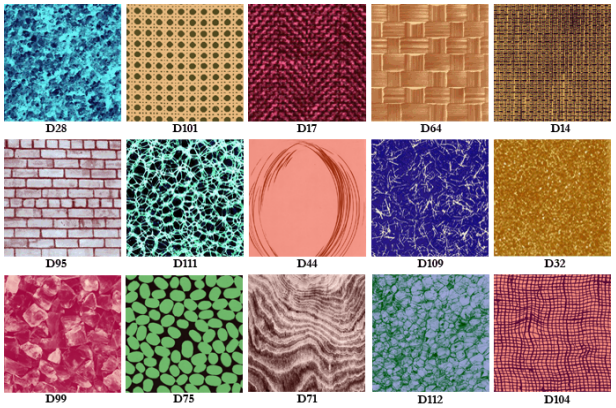


Fig. 6. Some texture classes from the color Brodatz (CBT) database [39].

112 textures of size 640×640 pixels. Each one is divided into 25 non-overlapping subimages of size 128×128 pixels, thus creating 2800 images in total (i.e. 112 classes $\times$ 25 images/class). The database is called CBT-2800 and some of its texture samples are displayed in Fig. 6.

B. Evaluation Criteria

The main criterion used to assess the performance of the proposed retrieval framework compared to state-of-the-art methods is the average retrieval rate (ARR). Let N_t, N_R be the total number of images in the database and the number of relevant images for each query, and for each query image q, let $n_q(K)$ be the number of correctly retrieved images among the K retrieved ones (i.e. K best matches). ARR in terms of number of retrieved images (K) is given by:

$$\mathrm{ARR}(K) = \frac{1}{N_t \times N_R} \sum_{q=1}^{N_t} n_q(K) \Big|_{K \geq N_R} \tag{16}$$

We note that K is generally set to be greater than or equal to N_R. By setting K equal to N_R, ARR becomes the primary benchmark considered by most studies to evaluate and compare the performance of different CBIR systems. All of ARR results shown in this paper are adopted by setting $K = N_R$. Last but not least, for better following our experimental study, Table I resumes the values of N_t and N_R, as well as the number of classes (N_c), of our three tested image databases.

TABLE I
NUMBER OF TOTAL IMAGES (N_t), NUMBER OF CLASSES (N_c) AND NUMBER OF RELEVANT IMAGES (N_R) FOR EACH QUERY WITHIN THE THREE TESTED DATABASES

	Vistex-640	**Stex-7616**	**CBT-2800**
N_t	640	7616	2800
N_c	40	476	112
N_R	16	16	25

C. Results and Discussion

1) Performance in retrieval accuracy: The proposed CBIR system has been performed on the three studied databases.

TABLE II
AVERAGE RETRIEVAL RATE (ARR) ON **VISTEX-640** DATABASE BY THE PROPOSED METHOD COMPARED TO THE STATE-OF-THE-ART METHODS

Method	**Using color**	**ARR** (%)
2002 GT+GGD+KLD [5]	-	76.57
2005 DT-CWT [6]	-	80.78
2005 DT-CWT+DT-RCWT [6]	-	82.34
2008 MGG+Gaussian+KLD [13]	✓	87.40
2008 MGG+Laplace+GD [13]	✓	91.70
2010 DCT+MGMM [8]	-	84.94
2011 Gaussian Copula+Gamma+ML [14]	✓	89.10
2011 Gaussian Copula+Weibull+ML [14]	✓	89.50
2011 Student-t Copula+GG+ML [14]	✓	88.90
2012 LMEBP [17]	-	87.77
2012 Gabor LMEBP [17]	-	87.93
2012 LtrP [19]	-	90.02
2012 Gabor LtrP [19]	-	90.16
2013 LEP+colorhist [20]	✓	82.65
2013 MCMCM+DBPSP [40]	✓	86.17
2014 Gaussian Copula-MWbl [10]	-	84.41
2015 NWST+WD [12]	-	85.30
2015 ODBTC [27]	✓	90.67
2015 EDBTC [28]	✓	92.55
2015 DDBTC [29]	✓	92.65
2015 LECoP [22]	✓	92.99
Proposed LED+RD (27-D)	✓	**94.64**
Proposed LED+RD (33-D)	✓	**94.70**

TABLE III
PER-CLASS RETRIEVAL RATE (%) ON THE **VISTEX-640** DATABASE USING THE PROPOSED LED+RD METHOD

Class	**27-D**	**33-D**	**Class**	**27-D**	**33-D**
Bark.0000	76.95	75.00	Food.0008	100.00	100.00
Bark.0006	98.05	98.05	Grass.0001	94.53	93.36
Bark.0008	83.20	84.38	Leaves.0008	99.61	100.00
Bark.0009	77.13	78.13	Leaves.0010	100.00	100.00
Brick.0001	100.00	99.22	Leaves.0011	100.00	100.00
Brick.0004	97.66	98.05	Leaves.0012	55.86	56.25
Brick.0005	100.00	100.00	Leaves.0016	86.72	86.72
Buildings.0009	100.00	100.00	Metal.0000	98.83	99.61
Fabric.0000	100.00	100.00	Metal.0002	100.00	100.00
Fabric.0004	76.95	77.34	Misc.0002	100.00	100.00
Fabric.0007	99.61	99.61	Sand.0000	100.00	100.00
Fabric.0009	100.00	100.00	Stone.0001	82.42	82.42
Fabric.0011	100.00	100.00	Stone.0004	90.63	91.02
Fabric.0014	100.00	100.00	Terrain.0010	94.92	95.31
Fabric.0015	100.00	100.00	Tile.0001	91.80	89.45
Fabric.0017	96.48	97.66	Tile.0004	100.00	100.00
Fabric.0018	98.83	100.00	Tile.0007	100.00	100.00
Flowers.0005	100.00	100.00	Water.0005	100.00	100.00
Food.0000	100.00	100.00	Wood.0001	96.88	96.88
Food.0005	99.22	99.22	Wood.0002	88.28	90.23
			ARR	**94.64**	**94.70**

For all cases, the local maximum and local minimum sets (i.e. $S_{\omega_2}^{\max}$ and $S_{\omega_2}^{\min}$) are extracted using a 3×3 search window ($\omega_2 = 3$). The local max keypoints are detected by the 5×5 window ($\omega_1 = 5$) and the neighborhood size for LED descriptor construction (W) is set to 30×30 pixels. Here, we set $\omega_1 > \omega_2$ to accelerate the computational time. Meanwhile, as discussed in Section III-A, a denser or coarser density of keypoints (by varying ω_1) can produce stable performance. Our experiments show that ω_1 set from 3 to 9 yields very close

ARR. Similarly, a value of W from 15 to 40 can provide quite equivalent retrieval results. We deeply discuss the sensitivity of our algorithm to these parameters later in Section IV-C3.

TABLE IV
AVERAGE RETRIEVAL RATE (ARR) ON **STEX-7616** DATABASE BY THE PROPOSED METHOD COMPARED TO THE STATE-OF-THE-ART METHODS

Method	**Using color**	**ARR** (%)
2002 GT+GGD+KLD [5]	-	*49.30
2008 DT-CWT+Weibull+KLD [7]	-	*58.80
2008 MGG+Laplace+GD [13]	✓	*71.30
2010 DWT+Gamma+KLD [9]	-	*52.90
2011 Gaussian Copula+Gamma+ML [14]	✓	69.40
2011 Gaussian Copula+Weibull+ML [14]	✓	70.60
2011 Student-t Copula+GG+ML [14]	✓	65.60
2013 LEP+colorhist [20]	✓	59.90
2015 DDBTC [29]	✓	44.79
2015 LECoP [22]	✓	74.15
Proposed LED+RD (27-D)	✓	**79.95**
Proposed LED+RD (33-D)	✓	**80.08**

(*) These results are extracted from [14], not from the original papers.

TABLE V
AVERAGE RETRIEVAL RATE (ARR) ON **CBT-2800** DATABASE BY THE PROPOSED METHOD COMPARED TO THE STATE-OF-THE-ART METHODS

Method	**Using color**	**ARR** (%)
2002 LBP [15]	-	*81.75
2012 LtrP [19]	-	*82.05
2014 LOCTP-YCbCr [21]	✓	84.46
2014 LOCTP-HSV [21]	✓	88.60
2014 LOCTP-LAB [21]	✓	88.90
2014 LOCTP-RGB [21]	✓	93.89
Proposed LED+RD (27-D)	✓	**99.06**
Proposed LED+RD (33-D)	✓	**98.79**

(*) These results are extracted from [21], not from the original papers.

Tables II, IV and V show the average retrieval rate (ARR) of the proposed framework performed on the Vistex-640, Stex-7616 and CBT-2800 databases, respectively, compared to several state-of-the-art methods in the literature. From these tables, the first observation is that recent CBIR propositions have taken into consideration more and more color information in order to improve their retrieval performance. This trend is reasonable since color components play a significant role in natural images. The second remark is that, as previously mentioned in the introduction of our article, the local patterns-based schemes (such as LtrP [19], LECoP [22], etc.) and the OTB-based systems [27]–[29] generally provide higher ARR than the wavelet-based probabilistic approaches [5]–[14]. Then, more importantly, our LED+RD framework (both the 27-D version and the improved 33-D version) has outperformed all reference methods for all the three databases. We now discuss the results for each database to validate the effectiveness of the proposed strategy.

An improvement in ARR of 1.71% (i.e. from 92.99% to 94.70%) and 5.93% (i.e. from 74.15% to 80.08%) has been achieved for the Vistex-640 and Stex-7616 databases, respectively, using the proposed algorithm compared to the local extrema co-occurrence pattern (LECoP) [22], the most efficient state-of-the-art method (in terms of ARR) in Tables II and IV. Within the proposed strategy, the 33-D improved LED descriptor gives a slightly higher ARR compared to the 27-D version (i.e. 0.06% for Vistex-640 and 0.13% for Stex-7616). This means that our framework could be still enhanced by investigating other features to modify and add into LED descriptors. Next, an important remark is that during our experimentation, most of the texture classes with strong structures and local features such as buildings, fabric categories, man-made object's surfaces, etc. are retrieved with 100% accuracy. Table III shows the per-class retrieval rate for each image class in the Vistex-640 database. As observed from the table, half of all the classes (19/40 classes by 27-D LED and 20/40 classes by 33-D LED) have been indexed with 100% accuracy. These perfectly-retrieved image classes in general consist of many local textures and structures. Similar behavior has been remarked for the Stex-7616 database. This issue is quite obvious since our motivation and expectation in this work is to exploit and emphasize local features inside each image in order to tackle the retrieval problem.

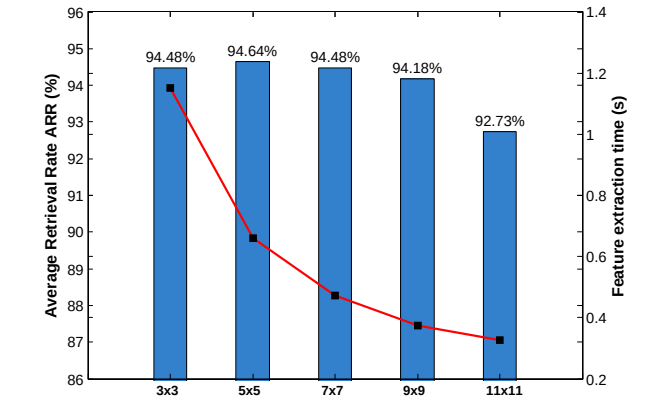
For the CBT-2800 database, a nearly perfect retrieval performance has been delivered by the proposed method. In Table V, the 27-D LED version yields slightly better performance (99.06%) than the 33-D version (98.79%). Compared to local oppugnant color texture pattern (LOCTP) [21], an enhancement of 5.17% in ARR (i.e. 99.06% compared to 93.89%) has been made. Since the CBT database has been recently created [39], there are not many studies working on it. In [21], the authors only considered 110 texture classes to create a database of 2750 sub-images, instead of 2800 sub-images from 112 classes as in our work (ref. Section IV-A). However, since our database is more complete and ARR is supposed not to change too much within 110 or 112 classes, we still consider the retrieval performance in [21] as reference results for our comparative study in Table V. To this end, in terms of retrieval accuracy, the efficiency of the proposed CBIR framework compared to reference methods is confirmed and validated for all tested databases.

TABLE VI
PERFORMANCE OF THE PROPOSED METHOD IN TERMS OF FEATURE EXTRACTION (FE) TIME AND DISSIMILARITY MEASUREMENT (DM) TIME. EXPERIMENTS ARE PERFORMED ON THE **VISTEX-640** DATABASE.

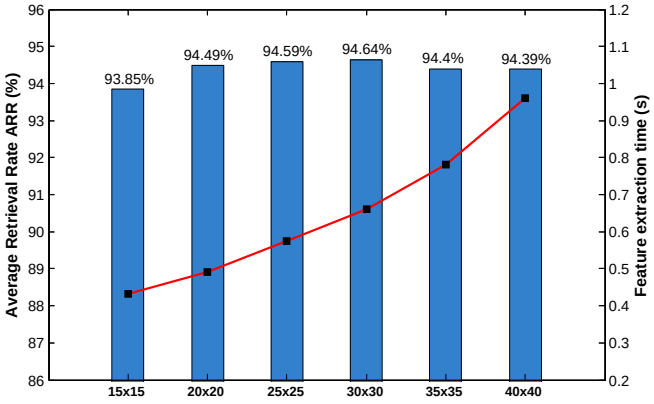
Version	**FE time** (s)		**DM time** (s)		**Total time** (s)		**ARR** (%)
	t_{data}	t_{image}	t_{data}	t_{image}	t_{data}	t_{image}	
27-D	422.8	0.661	22.3	0.035	445.1	0.695	94.64
33-D	476.6	0.745	28.1	0.044	504.7	0.789	94.70

t_{data}: time for the total database (640 images); t_{image}: time for each image.

2) Computation time: Table VI displays the computational cost required by the proposed strategy in terms of feature extraction (FE) time and dissimilarity measurement (DM) time performed on the Vistex-640 database. All implementations have been carried out using MATLAB 2012 on a laptop machine *Core i7-3740QM 2.7GHz, 16GB RAM*. In short, a total time of 445.1 seconds is required by the 27-D version to yield an ARR of 94.64% for the Vistex-640 database. For the 33-D version, it needs 504.7 seconds to deliver an accuracy



(a) Sensitivity to window size ω_1 for keypoint extraction



(b) Sensitivity to window size W for descriptor generation

Fig. 7. Sensitivity of the proposed method to its parameters in terms of average retrieval rate (%) and feature extraction time (s). Experiments are performed on Vistex-640 data set using the 27-D LED+RD.

of 94.70%. One remark is that the DM stage only requires 0.035s (resp. 0.044s) to measure the distance from each query image to 640 candidates from the database. In fact, to compute the riemannian distance in (15), only a generalized eigenvalue problem with matrix size of 27×27 (resp. 33×33) needs to be solved. Our experiments proved that this distance is less costly than other metrics (i.e. except the log-euclidean) mentioned in Section III-B. A detailed comparison will be given later in Section IV-C4.

3) Sensitivity analysis: This subsection aims at studying the sensitivity of the proposed method to its parameters. As observed from the full algorithm in Section III-C, only three parameters are required to perform our retrieval system: the window size to detect keypoints (ω_1), the window size to extract local maximum and local minimum pixels (ω_2) and the window size to generate LED descriptors (W). Since ω_2 needs to be small enough to ensure a sufficiently good density of local extrema to support the computation of LED descriptor. It is fixed to 3×3 pixels in our work. We now investigate the sensitivity to the other two parameters, ω_1 and W.

Fig. 7(a) shows the performance of the LED+RD algorithm (version 27-D) obtained by fixing $\omega_2 = 3$, $W = 30$ and varying ω_1 from 3 to 11. Experiments are performed on the Vistex-640 data. We observe a stable performance in ARR (varying from 94.18% to 94.64%) for $3 \leq \omega_1 \leq 9$. Then, a decrease of 1.45% in ARR is resulted when it switches from 9 to 11. In terms of computational cost, the feature extraction time is significantly reduced when ω_1 increased. Although the best performance in ARR (94.64%) is obtained by setting $\omega_1 = 5$, one may prefer setting it to 7 or 9 to speed up the computation time. Indeed, by using $\omega_1 = 7$, we gain about 28% of time (i.e. reduced from 0.661s to 0.473s per image) but only a reduction of 0.16% in ARR (i.e. from 94.64% to 94.48%) is resulted. Similarly, an ω_1 set to 9 could save 44% of time (i.e. reduced from 0.661s to 0.373s per image) and yields 94.18% retrieval accuracy (i.e. reducing only 0.46%).

Similarly, the algorithm sensitivity to W (i.e. window size used for the generation of LED descriptors) can be found in Fig. 7(b). Again, a stable performance can be observed when W varies from 15 to 40. Here, we set ω_1 to 5 and ω_2 always to 3. The highest ARR is adopted by setting W to 30. However, a smaller window size $W = 20$ may be more interesting since it reduces only 0.15% of ARR (i.e. reduced from 94.64% to 94.49%) while speeds up about 21% the feature extraction time (i.e. from 0.661s to 0.523s per image). As a result, a total time of only 349.2 seconds is necessary for our framework to produce 94.49% retrieval accuracy for Vistex-640 database. This issue makes the proposed strategy become very effective and competitive in both retrieval performance and computational cost. Furthermore, the two figures 7(a) and 7(b) show that the proposed LED method is not very sensitive to its parameters. A stable performance can be adopted with a wide range of parameters: $\omega_1 \in [3,\ 9]$ and $W \in [15,\ 40]$.

4) Sensitivity to distance measure: It is stated in Section III-B that the riemannian distance (RD) in (15) has outperformed other metrics we have implemented during the experimentation. In this subsection, we study and compare the performance of the proposed scheme by investigating different distance measures. The experiments have been done on the Vistex-640 database using the 27-D LED descriptors. Comparative results are shown in Table VII in terms of computational time and retrieval rate. From the table, the riemannian distance produces the highest ARR (i.e. 94.64%) and the second-lowest computational cost (i.e. 0.0349s, only greater than the log-euclidean metric with 0.0316s) to measure the distance from each query image to 640 candidates. The simplified Mahalanobis and the symmetric Kullback-Leibler distances (see (13) and (14)) consider that each point cloud follows a multivariate normal distribution with mean vector μ and covariance matrix C. They both yield interesting retrieval results (i.e. 90.48% and 91.82%, respectively). On the other hand, by not accounting for the mean feature vectors of those point clouds and considering the topological space of the feature covariance matrices (riemannian manifolds), some geometric-based metrics such as Wishart-like and the recommended riemannian could be more relevant. The Wishart-like distance in the second-last row gives a good performance in terms of ARR (92.39%) and requires 0.059s of computation time per image. However, the recommended riemannian distance shows better behaviors in both time consumption and retrieval rate. Besides that, two other metrics including the log-euclidean and the Bartlett distances seem to be not suitable with very low retrieval rate. Therefore, our choice of riemannian distance for

TABLE VII
SENSITIVITY TO DISTANCE MEASURES IN TERMS OF DISSIMILARITY MEASUREMENT TIME AND AVERAGE RETRIEVAL RATE (ARR). EXPERIMENTS ARE PERFORMED ON THE **VISTEX-640** DATABASE USING THE 27-D LED DESCRIPTORS.

Distance measure	**Formula**	t_{data} (s)	t_{image} (ms)	**ARR** (%)
Taking into account mean feature vectors				
Simplified Mahalanobis	$(\mu_1-\mu_2)^T\left(C_1^{-1}+C_2^{-1}\right)(\mu_1-\mu_2)$	40.51	63.30	90.48
Symmetric Kullback-Leibler	$\text{trace}\left(C_1C_2^{-1}+C_2C_1^{-1}\right)+(\mu_1-\mu_2)^T\left(C_1^{-1}+C_2^{-1}\right)(\mu_1-\mu_2)$	47.10	73.59	91.82
Not accounting for mean feature vectors				
Log-euclidean	$\|\|\log(C_1)-\log(C_2)\|\|$	**20.21**	**31.58**	72.65
Bartlett	$\log\frac{\lvert C_1+C_2\rvert^2}{\lvert C_1\rvert\lvert C_2\rvert}$	27.03	42.23	76.51
Wishart-like	$\text{trace}\left(C_1C_2^{-1}+C_2C_1^{-1}\right)$	37.79	59.90	92.39
Riemannian	$\sqrt{\sum_{\ell=1}^{d}\log^2\lambda_\ell}$, where $\lambda_\ell C_1\chi_\ell-C_2\chi_\ell=0, \ell=1\ldots d$	22.34	34.91	**94.64**

t_{data}: time for the entire database (640 images); t_{image}: time for each image; $||C||$: the Frobenius norm [36] of matrix C; $|C|$: the determinant of C.

dissimilarity measurement task is confirmed and validated.

V. CONCLUSION

A novel texture and color-based image retrieval framework has been proposed in this work by exploiting the local extrema-based features and riemannian metric. During the article, we have demonstrated the capacity of the local extrema pixels to capture significant information from the image content in order to encode color and local textural features. The proposed LED descriptor has proved its efficiency and relevance to tackle CBIR task. It is easy to implement, feasible to extend or improve, and not very sensitive to parameters. Then, by using such a riemannian distance to measure the dissimilarity between their feature covariance matrices, the proposed strategy has produced very competitive results in terms of both computational cost and average retrieval rate. It is expected to become one of the state-of-the-art methods in this field.

Future work can improve the performance of LED descriptor by exploiting other features within its generation. One may be interested by a deeper study on LED feature space to propose some other efficient distance metrics. It should be noted that the proposed framework can be applied to other types of images for retrieval and classification tasks within other application fields such as remote sensing imagery (i.e. optical, Radar, Sonar), medical imaging, etc. In these cases, it is recommended to investigate and adapt the feature extraction stage for the construction of LED descriptor.

VI. ACKNOWLEDGEMENTS

The authors would like to thank the French Space Agency (CNES) and the Brittany Region for funding this work. Also, we would like to thank Prof. Ronan FABLET for his fruitful discussion.

REFERENCES

[1] A. Alzubi, A. Amira, and N. Ramzan, "Semantic content-based image retrieval: A comprehensive study," *J. Visual Communi. Image Represent.*, vol. 32, pp. 20–54, 2015.

[2] T. Dharani and I. L. Aroquiaraj, "A survey on content based image retrieval," in *IEEE Int. Conf. Pattern Recogni. Inform. Mobile Engin. (PRIME)*, 2013, pp. 485–490.

[3] R. Veltkamp, H. Burkhardt, and H.-P. Kriegel, *State-of-the-art in content-based image and video retrieval*. Springer Science & Business Media, 2013, vol. 22.

[4] J. Ahmad, M. Sajjad, I. Mehmood, S. Rho, and S. W. Baik, "Describing colors, textures and shapes for content based image retrieval - A survey," *arXiv preprint arXiv:1502.07041*, 2015.

[5] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, 2002.

[6] M. Kokare, P. K. Biswas, and B. N. Chatterji, "Texture image retrieval using new rotated complex wavelet filters," *IEEE Trans. Syst. Man Cybern., Part B: Cybern.*, vol. 35, no. 6, pp. 1168–1178, 2005.

[7] R. Kwitt and A. Uhl, "Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2008, pp. 933–936.

[8] R. Kwitt and A. Uhl, "Lightweight probabilistic texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 241–253, 2010.

[9] S.-K. Choy and C.-S. Tong, "Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 281–289, 2010.

[10] N.-E. Lasmar and Y. Berthoumieu, "Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2246–2261, 2014.

[11] C. Li, G. Duan, and F. Zhong, "Rotation invariant texture retrieval considering the scale dependence of Gabor wavelet," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2344–2354, 2015.

[12] A. Sagel, D. Meyer, and H. Shen, "Texture retrieval via the scattering transform," *arXiv preprint arXiv:1501.02655*, 2015.

[13] G. Verdoolaege, S. De Backer, and P. Scheunders, "Multiscale colour texture retrieval using the geodesic distance between multivariate generalized Gaussian models," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2008, pp. 169–172.

[14] R. Kwitt, P. Meerwald, and A. Uhl, "Efficient texture image retrieval using copulas in a Bayesian framework," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2063–2077, 2011.

[15] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[16] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, 2010.

[17] M. Subrahmanyam, R. Maheshwari, and R. Balasubramanian, "Local maximum edge binary patterns: a new descriptor for image retrieval and object tracking," *Signal Process.*, vol. 92, no. 6, pp. 1467–1479, 2012.

[18] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.

[19] S. Murala, R. Maheshwari, and R. Balasubramanian, "Local tetra patterns: a new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, 2012.

[20] S. Murala, Q. J. Wu, R. Balasubramanian, and R. Maheshwari, "Joint histogram between color and local extrema patterns for object tracking," in *IS&T/SPIE Electronic Imaging*. Int. Soc. Optics Photonics, 2013, pp. 86 630T–86 630T–7.

[21] I. J. Jacob, K. Srinivasagan, and K. Jayapriya, "Local oppugnant color texture pattern for image retrieval system," *Pattern Recogn. Letters*, vol. 42, pp. 72–78, 2014.

[22] M. Verma, B. Raman, and S. Murala, "Local extrema co-occurrence pattern for color and texture image retrieval," *Neurocomputing*, vol. 165, pp. 255–269, 2015.

[23] G. Qiu, "Color image indexing using BTC," *IEEE Trans. Image Process.*, vol. 12, no. 1, pp. 93–101, 2003.

[24] M. R. Gahroudi and M. R. Sarshar, "Image retrieval based on texture and color method in BTC-VQ compressed domain," in *Proc. 9th IEEE Int. Symp. Signal Process. Applicat.*, 2007, pp. 1–4.

[25] F.-X. Yu, H. Luo, and Z.-M. Lu, "Colour image retrieval using pattern co-occurrence matrices based on BTC and VQ," *Electronics Lett.*, vol. 47, no. 2, pp. 100–101, 2011.

[26] J.-M. Guo, H. Prasetyo, and H.-S. Su, "Image indexing using the color and bit pattern feature fusion," *J. Vis. Commun. Image Repres.*, vol. 24, no. 8, pp. 1360–1379, 2013.

[27] J.-M. Guo and H. Prasetyo, "Content-based image retrieval using features extracted from halftoning-based block truncation coding," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1010–1024, 2015.

[28] J.-M. Guo, H. Prasetyo, and J.-H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," *IEEE Trans. Circuits Syst. Video Technol*, vol. 25, no. 3, pp. 466–481, 2015.

[29] J.-M. Guo, H. Prasetyo, and N.-J. Wang, "Effective image retrieval system using dot-diffused block truncation coding features," *IEEE Trans. Multimedia*, vol. 17, no. 9, pp. 1576–1590, 2015.

[30] M.-T. Pham, G. Mercier, and J. Michel, "Pointwise graph-based local texture characterization for very high resolution multispectral image classification," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 8, no. 5, pp. 1962–1973, 2015.

[31] M.-T. Pham, G. Mercier, and J. Michel, "Wavelets on graphs for very high resolution multispectral image segmentation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2014, pp. 2273–2276.

[32] W. Förstner and B. Moonen, "A metric for covariance matrices," in *Geodesy-The Challenge of the 3rd Millennium*. Springer, 2003, pp. 299–309.

[33] "Vision texture," MIT Vision and Modeling group, Available online: http://vismod.media.mit.edu/pub/VisTex/.

[34] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[35] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley and Sons, Ltd, 2000.

[36] I. L. Dryden, A. Koloydenko, and D. Zhou, "Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging," *The Annals of Applied Statistics*, pp. 1102–1123, 2009.

[37] A. C. Frery, A. D. Nascimento, and R. J. Cintra, "Analytic expressions for stochastic distances between relaxed complex wishart distributions," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1213–1226, 2014.

[38] R. Kwitt and P. Meerwald, "Salzburg texture image database," Available online: http://www.wavelab.at/sources/STex/.

[39] S. Abdelmounaime and H. Dong-Chen, "New Brodatz-based image databases for grayscale color and multiband texture analysis," *ISRN Machine Vision*, vol. 2013, 2013.

[40] M. Subrahmanyam, Q. J. Wu, R. Maheshwari, and R. Balasubramanian, "Modified color motif co-occurrence matrix for image indexing and retrieval," *Comput. Elect. Engin.*, vol. 39, no. 3, pp. 762–774, 2013.